



图文 128 案例实战：大型电商网站的上亿数据量的用户表如何进行水平拆分？

📱 手机观看

276 人次阅读 2020-11-11 07:00:00

详情 目录 评论

案例实战：大型电商网站的上亿数据量的用户表如何进行水平拆分？

C2C 电商系统微服务架构 120 天实战训练营

基于业界最流行的
Spring Cloud Alibaba 进行讲授

课程长度：84讲 价格：9块9

石杉老师最新出品：《C2C 电商系统微服务架构 120 天实战训练营》，点击了解详情

儒猿学院官网上线，内有石杉老师架构课最新大纲，儒猿云平台详细介绍，敬请浏览

官网：www.ruyuan2020.com（建议PC端访问）

今天开始会用最后两周的时间给大家讲解一下数据库分库分表的实战案例，首先，关于分库分表需要大家有一些基础知识的掌握和积累，关于分库分表的基础知识，请大家去看一下儒猿技术窝里的《互联网Java工程师面试突击第一季》，那是完全免费的一个课程，里面有关于分库分表的一些基础知识。



儒猿技术窝

进店逛逛

相关频道



从0开始带你成为MySQL
实战优化高手
已更新139期

其次，我们专栏里仅仅是讲解分库分表的整体方案的设计，但是在进行具体的方案落地的时候，是需要数据库中间件技术的支持的，业内常用的一般有Sharding-Sphere以及MyCat两种，各自用的公司都很多，都比较成熟，大家可以自行选择一个数据库中间件技术，去关注一下他们的官方文档，熟悉一下他们的用法。

因为上面介绍的两个数据库中间件技术都是国内开源的，所以文档都是有中文版的，大家阅读起来也是完全不费劲的。当大家有了上述的分库分表技术积累之后，那么就可以直接来阅读我们专栏里的分库分表案例实践了。

今天要给大家介绍的是平时我们见到的互联网公司里的海量用户数据的分库分表的方案，其实任何一个互联网公司都会有用户中心，这个用户中心就是负责这家公司的所有用户的数据管理，包括了用户的数据存储，用户信息的增删改查，用户的注册登录之类的。

而且一般互联网公司因为是直接面向终端用户的，所以用户数据一般都很多，哪怕是一个小互联网公司都能轻松拥有几百万级别的用户，如果是中型的互联网公司，一般达到几千万级别的用户都没问题，如果是大型互联网公司呢？那就是上亿，甚至几个亿级别的用户，甚至如果是BAT一类的巨头，用户体量可能是十亿级的。

所以说互联网公司的用户数据一般都是极为庞大的，那么我们今天的案例就选择一个背景是一个中型的电商公司，不是那种顶级电商巨头，就算是一个垂直领域的中型电商公司吧，那么他覆盖的用户大概算他有1亿以内，大概几千万的样子，就以这么个公司背景和用户量级来开展我们的案例。

首先，大家要明确一个要点，就是一般面对这么一个几千万级的数据，刚开始可能都是把数据放在MySQL的一个单库单表里的，但是往往这么大量级的数据到了后期，会搞的数据库查询速度很慢，因为结合之前的知识大家知道，数据量级太大了，会导致表的索引很大，树的层级很高，进而导致搜索性能下降，而且能放内存缓存的数据页是比较少的。

所以说，往往我们都建议MySQL单表数据量不要超过1000万，最好是在500万以内，如果能控制在100万以内，那是最佳的选择了，基本单表100万以内的数据，性能上不会有太大的问题，前提是，只要你建好索引就行，其实保证MySQL高性能通常没什么特别高深的技巧，就是控制数据量不要太大，另外就是保证你的查询用上了索引，一般就没问题。

好，所以针对这个问题，我们就可以进行分库分表了，可以选择把这个用户大表拆分为比如100张表，那么此时几千万数据瞬间分散到100个表里去，类似user_001、user_002、user_100这样的100个表，每个表也就几十万数据而已。

其次，可以把这100个表分散到多台数据库服务器上去，此时要分散到几台服务器呢？你要考虑两个点，一个是数据量有多少个GB/TB，一个是针对用户中心的并发压力有多高。实际上一般互联网公司对用户中心的压力不会高的太离谱，因为一般不会有很多人同时注册/登录，或者是同时修改自己的个人信息，所以并发这块不是太大问题。

至于数据量层面的话，我可以给大家一个经验值，一般1亿行数据，大致在1GB到几个GB之间的范围，这个跟具体你

一行数据有多少字段也有关系，大致大致就是这么个范围，所以说你几千万的用户数据，往多了说也就几个GB而已。

这点数据量，对于服务器的存储空间来说，完全没压力，不是问题。

所以综上所述，此时你完全可以给他分配两台数据库服务器，放两个库，然后100张表均匀分散在2台服务器上就可以了，分的时候需要指定一个字段来分，一般来说会指定userid，根据用户id进行hash后，对表进行取模，路由到一个表里去，这样可以让数据均匀分散。

到此就搞定了用户表的分库分表，你只要给系统加上数据库中间件技术，设置好路由规则，就可以轻松的对2个分库上的100张表进行增删改查的操作了。平时针对某个用户增删改查，直接对他的userid进行hash，然后对表取模，做一个路由，就知道到哪个表里去找这个用户的数据了。

但是这里可能会出现一些问题，一个是说，用户在登录的时候，可能不是根据userid登陆的，可能是根据username之类的用户名，手机号之类的来登录的，此时你又没有userid，怎么知道去哪个表里找这个用户的数据判断是否能登录呢？

关于这个问题，一般来说常规方案是建立一个索引映射表，就是说搞一个表结构为（username, userid）的索引映射表，把username和userid——映射，然后针对username再做一次分库分表，把这个索引映射表可以拆分为比如100个表分散在两台服务器里。

然后用户登录的时候，就可以根据username先去索引映射表里查找对应的userid，比如对username进行hash然后取模路由到一个表里去，找到username对应的userid，接着根据userid进行hash再取模，然后路由到按照userid分库分表的一个表里去，找到用户的完整数据即可。

但是这种方式会把一次查询转化为两个表的两次查询，先查索引映射表，再根据userid去查具体的数据，性能上是有一定的损耗的，不过有时候为了解决分库分表的问题，也只能用这种类似的办法。

另外就是如果在公司运营团队里，有一个用户管理模块，需要对公司的用户按照手机号、住址、年龄、性别、职业等各种条件进行极为复杂的搜索，这怎么办呢？其实没太多的好办法，基本上就是要对你的用户数据表进行binlog监听，把你要搜索的所有字段同步到Elasticsearch里去，建立好搜索的索引。

然后你的运营系统就可以通过Elasticsearch去进行复杂的多条件搜索，ES是适合干这个事儿的，然后定位到一批userid，通过userid回到分库分表环境里去找出具体的用户数据，在页面上展示出来即可。

这就是一套比较常规和完整的分库分表的方案。

End


专栏版权归公众号儒猿技术窝所有

未经许可不得传播，如有侵权将追究法律责任

儒猿技术窝精品专栏及课程推荐：

- [《从零开始带你成为消息中间件实战高手》](#)
- [《互联网Java工程师面试突击》（第2季）](#)
- [《互联网Java工程师面试突击》（第1季）](#)
- [《互联网Java工程师面试突击》（第3季）](#)
- [《从零开始带你成为JVM实战高手》](#)
- [《C2C电商系统微服务架构120天实战训练营》](#)

Copyright © 2015-2020 深圳小鹅网络技术有限公司 All Rights Reserved. 粤ICP备15020529号

 小鹅通提供技术支持